METHOD FOR EFFICIENTLY STORING THE TRAJECTORY OF TRACKED OBJECTS IN VIDEO

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to the tracking of objects in video sequences. More particularly, the present invention relates to storage of coordinates used to track object trajectories.

2. Description of the Related Art

In the prior art, when objects are tracked in a video sequence, trajectory coordinates are typically generated for each frame of video. Considering that, for example, that under the NTSC standard, which generates 30 frames per second, a new location or coordinate for each object in a video sequence must be generated and stored for each frame.

This process is extremely inefficient and requires tremendous amounts of storage. For example, if five objects in a video sequence were tracked, over two megabytes of storage would be needed just to store the trajectory data for a single

hour. Thus, storage of all of the trajectories is expensive, if not impractical.

There have been attempts to overcome the inefficiency of the prior art. For example, in order to save space, the coordinates for every video frame have been compressed. One drawback is that the compression of the trajectories introduces delay into the process. Regardless of the compression, there is still a generation of coordinates for each frame. In addition, there has been an attempt to circumvent the generation of trajectories by devices that store the location of motion in video for every frame, based on a grid-based breakup of the video frame. These devices still store data for each frame, and the accuracy of the location of motion is not comparable to the generation of trajectories.

SUMMARY OF THE INVENTION

Accordingly, it is an object of the present invention to provide a method and system that addresses the shortcomings of the prior art.

In a first aspect of the present invention, the coordinates are stored only when objects move more than a predetermined

2

amount, rather than storing their movement after every frame.

This feature permits a tremendous savings in memory or disk usage over conventional methods. In addition, the need to generate coordinates can be greatly reduced to fractions of the generation per frame that is conventionally processed.

A video content analysis module automatically identifies objects in a video frame, and determines the (x_i, y_i) coordinates of each object i. The reference coordinates for each for object i, $(xref_i, yref_i)$ are set to (x_i, y_i) when the object is first identified. For subsequent frames, if the new coordinates $(xnew_i, ynew_i)$ are less than a given distance from the reference coordinates, that is if $\| (xnew_i, ynew_i) - (xref_i, yref_i) \|^2 < \varepsilon$, then the current coordinates are ignored. However, if the object moves more than the distance ε, the current coordinates $(xnew_i, ynew_i)$ are stored in the object's trajectory list, and we set the reference coordinates $(xref_i, yref_i)$ to the object's current position. This process is repeated for all subsequent video frames. The resulting compact trajectory lists can then be written to memory or disk while they are being generated, or when they are complete.

The present invention can be used in many areas, including video surveillance security system that tracks movement in a

particular area, such as a shopping mall, etc. The amount of storage conventionally required for standard video cameras that scan/videotape an area, such a VCR, often creates a huge unwanted library of tapes. In addition, there is a tendency to

5 reuse the tapes quickly so as not to set aside tape storage areas, or pay for their shipment elsewhere. The compact storage of the present invention makes the permanent storage of secure areas much more practical, and provides a record to investigators to see whether a particular place was "cased"

10 (e.g. observed by a wrongdoer prior to committing an unlawful act) by a wrongdoer prior to a subsequent unlawful action being performed.

Also, in a commercial setting, the present invention could

15 be applied to track people in, for example, a retail store to see how long they waited on the checkout line.

Accordingly, a method for storing a trajectory of tracked objects in a video, comprising the steps of:

20   (a) identifying objects in a first video frame;

        (b) determining first reference coordinates

$(xref_i, yref_i)$ for each of said objects identified in step (a) in the first video frame;

(c) storing the first reference coordinates $(xref_i, yref_i)$;

(d) identifying said objects in a second video frame;

(e) determining current reference coordinates

5 $(xnew_i, ynew_i)$ of said objects in said second video frame; and

(f) storing the current reference coordinates of a particular object in an object trajectory list and replacing the first reference coordinates $(xref_i, yref_i)$ with the current reference coordinates $(xnew_i, ynew_i)$ if the following condition

10 for the particular object is satisfied:

$$\| (xnew_i, ynew_i) - (xref_i, yref_i) \|^2 \geq \varepsilon,$$

wherein ε is a predetermined threshold amount, and

retaining the first reference coordinates $(xref_i, yref_i)$ for comparison with subsequent video frames when said condition in

15 step (f) is not satisfied.

The method according may further comprise (g) repeating steps (e) and (f) for all video frames subsequent to said second video frame in a video sequence so as to update the storage area

20 with additional coordinates and to update the current reference coordinates with new values each time said condition in step (f) is satisfied.

Optionally, the method may include a step of storing the last coordinates of the object (i.e., the coordinates just

5

before the object disappears and the trajectory ends), even if the last coordinate does not satisfy condition (f).

The object trajectory list for the particular object stored in step (f) may comprise a temporary memory of a processor, and the method may optionally include the folowing step:

(h) writing the object trajectory list to permanent storage from all the coordinates stored in the temporary memory after all the frames of the video sequence have been processed by steps (a) to (g).

The permanent storage referred to in step (h) may comprise at least one of a magnetic disk, optical disk, and magneto-optical disk, or even tape. Alternatively, the permanent storage can be arranged in a network server.

The determination of the current reference coordinates $(x_{1new}y_{inew})$ in step (e) can include size tracking of the objects moving one of (i) substantially directly toward, and (ii) substantially directly away from a camera by using a box bounding technique. The box bounding technique may comprise:

(i) determining a reference bounding box $(wref_i, href_i)$ of the particular object i, wherein w represents a width, and h represents a height of the particular object;

6

(ii) storing a current bounding box (w_i, h_i) if either of the following conditions in substeps (ii) (a) and (ii) (b) are satisfied:

(ii) (a) $\left| w_i - wref_i \right| > \delta_w$;

(ii) (b) $\left| h_i - href_i \right| > \delta_h,$

where δ_w and δ_h are predetermined thresholds.

Alternatively, the box bounding technique may comprise:

(i)   determining an area $aref_i = wref_i * href_i$ of a reference bounding box $(wref_i, href_i)$ of the particular object, wherein w represents a width, and h represents a height of the particular object; and

(ii) storing coordinates of a current bounding box (w_i, h_i) if a change in area $\delta_a = \left| aref_I - w_i * h_i \right|$ of the current bounding box is greater than a predetermined amount.

Brief Description of the Drawings

Figs. 1A-1C illustrate a first aspect of the present invention wherein the motion in Fig. 1B relative to Fig. 1A fails to satisfy the expression in Fig. 1C.

Figs. 2A-2C illustrate a second aspect of the present invention wherein the motion in Fig. 2B relative to Fig. 2A satisfies the expression in Fig. 1C.

Figs. 3A-3C illustrate another aspect of the present invention pertaining to a box bounding technique.

Fig. 4 illustrates a schematic of a system used according to the present invention.

Fig. 5A and 5B are a flow chart; illustrating an aspect of the present invention.

Detailed Description of the Preferred Embodiments

Figure 1A-1C illustrate a first aspect of the present invention. As shown in Figure 1A a frame 105 contains an object 100 (in this case a stick figure representing a person). To aid in understanding, numerical scales in both the X direction and Y direction have been added to the frame. It is noted that the x,y coordinates can be obtained, for example, by using the center of the mass of the object pixels, or in the case of a bounding box technique (which is disclosed, *infra*) by using the center of the object bounding box.

It should be understood by persons of ordinary skill in the art that the scales are merely for illustrative purposes, and the spaces there between, and/or the number values do not limit the claimed invention to the scale. The object 100 is

identified at a position ($xref_i$, $yref_i$) which are now used as the x and y reference point for this particular object.

It should be noted that the objects identified do not have to be, for example, persons, and could include inanimate objects in the room, such as tables, chairs, and desks. As known in the art, these objects could be identified by, for example, their color, shape, size, etc. Preferably, a background subtraction technique is used to separate moving objects from the background. One way this technique is used is by learning the appearance of the background scene and then identifying image pixels that differ from the learned background. Such pixels typically correspond to foreground objects. Applicants hereby incorporate by reference as background material the articles by A. Elgammal, D. Harwood, and L. Davis, "Non-parametric Model for Background Subtraction", *Proc. European Conf. on Computer vision*, pp. II: 751-767, 2000, and C. Stauffer, W.E.L. Grimson, "Adaptive Background Mixture Models for Real-time Tracking", *Proc. Computer Vision and Pattern Recognition*, pp. 246-252, 1999 as providing reference material for some of the methods that an artisan can provide object identification. In the Stauffer reference, simple tracking links objects in successive frames based on distance, by marking each object in the new frame by the same number as the closest object in the previous frame.

9

Additionally, the objects can be identified by grouping the foreground pixels, for example, by a conected-components algorithm, as described by T. Cormen, C. Leiserson, R. Rivest, "Introduction to Algorithms", MIT Press, 1990, chapter 22.1,

5 which is hereby incorporated by reference as background material. Finally, the objects can be tracked such as disclosed in U.S. patent application serial 09/xxx,xxx entitled "Computer Vision Method and System for Blob-Based Analysis Using a Probabilistic Network, U.S. serial 09/988,946 filed November

10 19, 2001, the contents of which are hereby incorporated by reference.

Alternatively, the objects could be identified manually. As shown in Figure 1B, object 100 has moved to a new position

15 captured in the second frame 110 having coordinates of $(xnew_i, ynew_i)$ which is a distance away from the $(xref_i, yref_i)$ of the first frame 105.

It is appreciated by an artisan that while there are many

20 ways that objects can be identified and tracked, the present invention is applicable regardless of the specific type of identification and tracking of the objects. The amount of savings in storage is significant irrespective of the type of identification and tracking.

According to an aspect of the present invention, rather than storing new coordinates for every object and every frame, an algorithm determines whether or not the movement by object 100 in the second frame is greater than a certain predetermined amount. In the case where the movement is less than the predetermined amount, coordinates for Figure 1B are not stored. The reference coordinates identified in the first frame 105 continue to be used against a subsequent frame.

Figure 2A again illustrates, (for convenience of the reader), frame 105, whose coordinates will be used to track motion in a third frame 210. The amount of movement by the object 100 in the third frame, as opposed to its position in the first frame 105, is greater than the predetermined threshold. Accordingly, the coordinates of the object 100 in Figure 2B now become the new reference coordinates (as identified in the drawing as new $(xref_i, yref_i)$, versus the old $(xref_i, yref_i)$. Accordingly, the trajectory of the object 100 includes the coordinates in frames 1 and 3, without the need to save the coordinates in frame 2. It should be understood that, for example, as standards such as NTSC generate 30 frames per second, the predetermined amount of movement could be set so that significant amounts of coordinates would not require

storage. This process can permit an efficiency in compression heretofore unknown.

The amount of movement used as a predetermined threshold could be tailored for specific applications, and includes that the threshold can be dynamically computed, or modified during the analysis process. The dynamic computation can be based on factors such as average object velocity, general size of the object, importance of the object, or other statistics of the video.

For example, in a security film, very small amounts of motion could be used when items being tracked are extremely valuable, as opposed to larger threshold amounts permit more efficient storage, which can be an important consideration based on storage capacity and/or cost. In addition, the threshold amount can be application specific so that the trajectory of coordinates is as close to the actual movement as desired. In other words, if a threshold amount is too large, it could be movement in different directions that is not stored. Accordingly, the trajectory of the motion would be that between only the saved coordinates, which, of course, may not necessarily comprise the exact path that would be determined in the conventional tracking and storage for each individual frame. It should be noted that with many forms of compression, there

12

normally is some degree of paring of the representation of the objects.

Figures 3A to 3C illustrate another aspect of the present invention pertaining to a box bounding technique. It is understood by persons of ordinary skill in the art that while a camera is depicted, the video image could be from a video server, DVD, videotape, etc. When objects move directly toward or away from a camera, their coordinates may not change enough to generate new trajectory coordinates for storage. A box bounding technique is one way that the problem can be overcome. For example, in the case of an object moving directly toward or away from the camera, the size of the object will appear to become larger or smaller depending on the relative direction.

Figures 3A to 3C illustrate a box bounding technique using size tracking. As shown in Figure 3A, a bounding box 305 represents the width and height of an object 307 the first frame 310.

As shown in the second frame 312 in Figure 3B, the bounding box in 310 of object 307 has changed (as these drawings are for explanatory purposes, they are not necessarily to scale).

13

As shown in Figure 3C, the box bounding technique would store the coordinate of the object in the second frame 312 if the width of a bounding box in a subsequent frame is different from the width of the reference box of the previous frame, or the height of the bounding box in a particular frame is different from the height of the bounding box of a reference frame; in each case the difference is more than a predetermined threshold value. Alternatively, the area of the bounding box (width x height) could be used as well, so if the area of the bounding box 310 is different than the area of the reference bounding box 305 by a predetermined amount, the coordinates of the second frame would be stored.

Figure 4 illustrates one embodiment of a system according to the present invention. It should be understood that the connections between all of the elements could be any combination of wired, wireless, fiber optic, etc. In addition, some of the items could be connected via a network, including but not limited to the Internet. As shown in Figure 4, a camera 405 captures images of a particular area and relays the information to a processor 410. The processor 410 includes a video content analysis module 415 which identifies objects in a video frame and determines the coordinates for each object. The current

14

reference coordinates for each object could be stored, for example, in a RAM 420, but it should be understood that other types of memory could be used. As a trajectory is a path, the initial reference coordinates of the identified objects would also be stored in a permanent storage area 425. This permanent storage area could be a magnetic disc, optical disc, magneto optical disc, diskette, tape, etc. or any other type of storage. This storage could be located in the same unit as the processor 410 or it could be stored remotely. The storage could in fact be part of or accessed by a server 430. Each time the video content module determines that motion for an object in a frame exceeds the value of the reference coordinates by a predetermined threshold, the current reference coordinates in the RAM 420 would be updated as well as permanently stored 425. As the system contemplates only a storage of motion beyond a certain threshold amount, the need to provide storage or sufficient capacity to record every frame is reduced and in most cases, eliminated. It should also be noted that the storage could be video tape.

Applicants' Figures 5A and 5B illustrate a flow chart that provides an overview of the process of the present of the present invention.

At step 500, objects in the first video frame are identified.

At step 510, the reference coordinates for each of the objects identified in the first video frame are determined. The determination of these reference coordinates may be known by any known method, e.g., using the center of the object bounding box, or the center of mass of the object pixels.

At step 520, the first reference coordinates determined in step 10 are stored. Typically, these coordinates could be stored in a permanent type of memory that would record the trajectory of the object. However, it should be understood that the coordinates need not be stored after each step. In other words, the coordinates could be tracked by the processor in the table, and after all the frames have been processed, the trajectory could be stored at that time.

At step 530, the objects in the second video frame are identified.

At step 540, there is a determination of the current reference coordinates of the objects in the second video frame. These coordinates may or may not be the same as in the first frame. As shown in Figure 5B, at step 550 the current reference coordinates of a particular object are stored in an object trajectory list and used to replace the first referenced coordinates of that particular object if the following condition

16

for the particular object is satisfied $\| (xnew_i, ynew_i) - (xref_i, yref_i) \|^2 \geq \varepsilon$, However, when the condition is not satisfied, the first reference coordinates are retained for comparison with subsequent video frames. The process continues

5 until all of the video frames have been exhausted. As previously discussed, the object trajectory list could be a table, and/or a temporary storage area in the processor which is later stored, for example, on a hard drive, writeable CD ROM, tape, non volatile electronic storage, etc. Various modifications may be

10 made on the present invention by a person of ordinary skill in the art that would not depart from the spirit of the invention or the scope of the appended claims. For example, the type of method used to identify the object in the video frames, the threshold values provided by which storage of additional

15 coordinates and subsequent frames is determined, may all be modified by the artisan in the spirit of the claimed invention. In addition, a time interval could be introduced into the process, where for example, after a predetermined amount of time, the coordinates of a particular frame are stored even if a

20 predetermined threshold of motion is not reached. Also, it is within the spirit of the invention and the scope of the appended claims, and understood by an artisan that that coordinates other than x and y could be used, (for example, z) or, the x,y coordinates could be transformed into another space, plane or

coordinate system, and the measure would be done in the new space. For example, if the images were put through a perspective transformation prior to measuring.   In additio, the distance measured could be other than Euclidian distance, such as a less-
5   compute-intensive measure, such as $|xnew-xref| + |ynew-yref| \geq$
ε.